SUBSTITUTE SPECIFICATION

Genes of the 1-deoxy-D-xylulose biosynthesis pathway

The present invention relates to DNA sequences which, when incorporated into the genome of viruses, eukaryotes and prokaryotes, modify isoprenoid biosynthesis and to a genetic engineering process for the production of these transgenic viruses, eukaryotes and prokaryotes. The invention also relates to a process for the identification of substances having herbicidal, antimicrobial, antiparasitic, antiviral, fungicidal, bactericidal action in plants and antimicrobial, antiparasitic, antimycotic, antibacterial and antiviral action in humans and animals.

The biosynthesis pathway for the formation of isoprenoids via the classical acetate/mevalonate pathway and an alternative mevalonate-independent biosynthesis pathway, the deoxy-D-xylulose phosphate pathway is already known (Rohmer, M., Knani, M., Simonin, P., Sutter, B. and Sahm, H. (1993): *Biochem. J.* 295: 517-524).

It is, however, not known how and by which pathways it is possible to bring about a change in the isoprenoid concentration in viruses, eukaryotes and prokaryotes by means of the deoxy-D-xylulose phosphate pathway. Figure 1 shows this biosynthesis pathway.

DNA sequences are consequently provided which code for 1-deoxy-D-xylulase 5-phosphate synthase (DOXP synthase), 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DOXP reductoisomerase) or the gcpE protein. All three genes and enzymes are involved in isoprenoid biosynthesis.

The gcpE protein has a kinase function and catalyses the phosphorylation of a sugar or a phosphorus sugar or a precursor of isoprenoid biosynthesis, in particular the phosphorylation of 2-C-methyl-D-erythritol, 2-C-methyl-D-erytritol phosphate, in particular 2-C-methyl-D-erythritol 4-phosphate, 2-C-methyl-D-erythrose, 2-C-methyl-D-erythrose phosphate, in particular 2-C-methyl-D-erythrose 4-phosphate. In the precursor of isoprenoid synthesis, the gcpE protein in particular catalyses the phosphorylation of the following substances:

$CH_2(OH)-C(CH_3)=C(OH)-CH_2-O-PO(OH)_2$,

$CH_2(OH)-C(CH_3)=C(OH)-CH_2-OH$,

$CH_2(OH)-CH(CH_3)-CO-CH_2-O-PO(OH)_2$,

$CH_2(OH)-CH(CH_3)-CO-CH_2OH$

$CH_2=C(CH_3)-CO-CH_2-O-PO(OH)_2$,

$CH_2=C(CH_3)-CO-CH_2-OH$,

$CH_2=C(CH_3)-CH(OH)-CH_2-O-PO(OH)_2$,

$CH_2=C(CH_3)-CH(OH)-CH_2-OH$,

$CH_2(OH)-C(=CH_2)-C(OH)-CH_2-O-PO(OH)_2$,

$CH_2(OH)-C(=CH_2)-C(OH)-CH_2-OH$

$CHO-CH(CH_3)-CH(OH)-CH_2-O-PO-(OH)_2$,

$CHO-CH(CH_3)-CH(OH)-CH_2-OH$,

$CH_2(OH)-C(OH)(CH_3)-CH=CH-O-PO(OH)_2$,

$CH_2(OH)-C(OH)(CH_3)-CH=CH-OH$

$CH(OH)=C(CH_3)-CH(OH)-CH_2-O-PO(OH)_2$,

$CH(OH)=C(CH_3)-CH(OH)-CH_2-OH$,

$(CH_3)_2HC-CO-CH_2-O-PO(OH)_2$,

$(CH_3)_2HC-CO-CH_2-O-H$,

$(CH_3)_2HC-CH(OH)-CH_2-O-PO(OH)_2$,

$(CH_3)_2HC-CH(OH)-CH_2-O-H$.

DOXP synthase catalyses the condensation of pyruvate and glyceraldehyde 3-phosphate to yield 1-deoxy-D-xylulose

5-phosphate and DOXP reductoisomerase catalyses the conversion of 1-deoxy-D-xylulose 5-phosphate into 2-C-methyl-D-erythritol 4-phosphate (*c.f.* Fig. 1).

The invention relates to the following DNA sequences: DNA sequences which code for a polypeptide with the amino acid sequence shown in SEQ ID no. 2 or for an analogue or derivative of the polypeptide according to SEQ ID no. 2, in which one or more amino acids have been deleted, added or replaced by other amino acids,

DNA sequences which code for a polypeptide with the amino acid sequence shown in SEQ ID no. 4 or for an analogue or derivative of the polypeptide according to SEQ ID no. 4, in which one or more amino acids have been deleted, added or replaced by other amino acids,

and DNA sequences which code for a polypeptide with the amino acid sequence shown in SEQ ID no. 6 or for an analogue or derivative of the polypeptide according to SEQ ID no. 6, in which one or more amino acids have been deleted, added or replaced by other amino acids.

The genes and the gene products thereof (polypeptides) are shown with their primary structure and are assigned as follows:

SEQ ID no. 1: 1-deoxy-D-xylulose 5-phosphate reducto-isomerase gene

SEQ ID no. 2: 1-deoxy-D-xylulose 5-phosphate reducto-isomerase

SEQ ID no. 3: 1-deoxy-D-xylulose 5-phosphate synthase gene

SEQ ID no. 4: 1-deoxy-D-xylulose 5-phosphate synthase

SEQ ID no. 5: gcpE gene

SEQ ID no. 6: gcpE proteins.

The DNA sequences all originate from *Plasmodium falciparum*.

Apart from the DNA sequences stated in the sequence listing, suitable sequences are also those which, as a result of the degeneration of the genetic code, have another DNA sequence, but code for the same peptide or for an analogue or derivative of the polypeptide, in which one or more amino acids have been deleted, added or replaced by other amino acids.

The sequences according to the invention are suitable for the expression of genes in viruses, eukaryotes and prokaryotes which are responsible for isoprenoid biosynthesis in the 1-deoxy-D-xylulose pathway.

According to the invention, eukaryotes or eukaryotic cells include animal cells, plant cells, algae, yeasts, fungi, while prokaryotes or prokaryotic cells include bacteria, archaebacteria and eubacteria.

When a DNA sequence is incorporated into a genome on which the above-stated DNA sequence is located, expression of the above-described genes in viruses, eukaryotes and prokaryotes is enabled. The viruses, eukaryotes and prokaryotes transformed according to the invention are cultivated in a manner known *per se* and the isoprenoid formed during such cultivation is isolated and optionally purified. Not all isoprenoids need to be

isolated as in some case the isoprenoids are released directly into the ambient air.

The invention furthermore relates to a process for the production of transgenic viruses, eukaryotes and prokaryotes in order to modify the isoprenoid content, which process comprises the following steps.

a) Production of a DNA sequence with the following sub-sequences
   i) promoter which is active in viruses, eukaryotes and prokaryotes and ensures the formation of an RNA in the intended target tissue or target cells,
   ii) DNA sequence which codes for a polypeptide with the amino acid sequence shown in SEQ ID no. 2, 4 or 6 or for an analogue or derivative of the polypeptide according to SEQ ID no. 2, 4 or 6,
   iii) 5' and 3' untranslated sequence which enables or enhances expression of the stated genes in viruses, eukaryotes and prokaryotes,
b) transfer and incorporation of the DNA sequence into the genome of viruses, prokaryotic or eukaryotic cells with or without the use of a vector (for example plasmid, viral DNA).

The intact, whole plants may be regenerated from plant cells transformed in this manner.

The protein-coding sequences with the nucleotide sequences SEQ ID no. 1, SEQ ID no. 3 and SEQ ID no. 5 may be provided with a promoter which ensures transcription in certain organs or cells, which promoter is coupled in

sense orientation (3' end of the promoter to the 5' end of the coding sequence) to the sequence which codes the protein to be formed. A termination signal which determines termination of mRNA synthesis is attached to the 3' end of the coding sequence. In order to direct the protein which is to be expressed to certain subcellular compartments, such as chloroplasts, amyloplasts, mitochondria, vacuoles, cytosol or intercellular spaces, a further sequence which codes for a so-called signal sequence or a transit peptide may be inserted between the promoter and the coding sequence. In some cases, it is necessary to insert sequences which code for a signal at the COOH terminus of the protein. The sequence must be in the same reading frame as the coding sequence of the protein. A large number of cloning vectors is available in order to prepare for the introduction of the DNA sequences according to the invention into higher plants, which vectors contain a replication signal for *E. coli* and a marker which permits selection of the transformed cells. Depending upon the method by which desired genes are introduced into the plant, further DNA sequences may be required. If, for example, the Ti or Ri plasmid is used to transform the plant cells, at least one right border, but frequently the right border and left border of the Ti and Ri plasmid T-DNA must be inserted as a flanking region into the genes to be introduced. The use of T-DNA for transforming plant cells has been intensively investigated and comprehensively described in EP 120516; Hoekama in "The Binary Plant Vector System", Offset-drukkerij Kanters B.V. Alblasserdam (1985), chapter V; Fraley *et al.*, *Crit.Rev.Plant Sci.* 4, 1-46 and An *et al.* (1985) *EMBO J.* 4, 277-287. Once the introduced DNA has been incorporated into the genome, it is

generally stable and is also retained in the descendants of the originally transformed cells. It normally contains a selection marker, which imparts to the transformed plant cells resistance to a biocide or an antibiotic, such as kanamycin, G 418, bleomycin, hygromycin or phosphinotricin and others. The particular marker used is thus intended to allow selection of transformed cells from cells lacking the inserted DNA.

Many techniques are available for introducing DNA into a plant. These techniques include transformation with the assistance of agrobacteria, for example *Agrobacterium tumefaciens*, protoplast fusion, microinjection of DNA, electroporation, as well as ballistic methods and virus infection. Whole plants may then be regenerated from the transformed plant material in a suitable medium which may contain antibiotics or biocides for selection purposes. No particular requirements are placed upon the plasmids for injection and electroporation. However, if whole plants are to be regenerated from such transformed cells, a selectable marker gene must be present. The transformed cells grow in the plants in the conventional manner (McCormick *et al.* (1986), *Plant Cell Reports* 5, 81-84). The plants may be cultivated normally and be crossed with plants which have the same transformed genome or other genomes. The resultant individuals have the corresponding phenotypic properties.

The present invention also provides expression vectors which contain one or more of the DNA sequences according to the invention. Such expression vectors are obtained by providing the DNA sequences according to the invention with suitable functional regulation signals. Such

regulation signals are DNA sequences which are responsible for expression, for example promoters, operators, enhancers, ribosomal binding sites, and are recognised by the host organism.

Further regulation signals, which for example control replication or recombination of the recombinant DNA in the host organism, may optionally also be a constituent part of the expression vector.

The host organisms transformed with the DNA sequences or expression vectors according to the invention are also provided by the present invention.

Suitable host cells and organisms for expressing the enzymes according to the invention are those which comprise no intrinsic enzymes with the function of DOXP synthase, DOXP reductoisomerase or the gcpE protein. This is the case for archaebacteria, animals, fungi, slime moulds and some eubacteria. The absence of such intrinsic enzyme activity substantially facilitates detection and purification of the recombinant enzymes. As a consequence, it is also for the first time possible straightforwardly to measure, in crude extracts from the host cells, the activity and in particular the inhibition of the activity of the recombinant enzymes according to the invention by various chemicals and pharmaceuticals.

The enzymes according to the invention are advantageously then expressed in eukaryotic cells if post-translational modification and native folding of the polypeptide chain is to be achieved. Moreover, depending upon the expression system, it is ensured when expressing genomic

DNA sequences that introns are eliminated by splicing the DNA and the enzymes are produced in the polypeptide sequences characteristic to the parasites. Using recombinant DNA techniques, sequences coding for introns may be eliminated from or inserted for experimental purposes into the DNA sequences to be expressed.

The protein may be isolated from the host cell or the culture supernatant of the host cell using methods known to the person skilled in the art. *In vitro* reactivation of the enzymes may also be required.

In order to facilitate purification, the enzymes according to the invention or sub-sequences of the enzymes may be expressed as fusion proteins with various peptide chains. Oligo-histidine sequences and sequences derived from glutathione S-transferase, thioredoxin or calmodulin-binding peptides are particularly suitable for this purpose.

The enzymes according to the invention or sub-sequences of the enzymes may furthermore be expressed as fusion proteins with such peptide chains known to the person skilled in the art that the recombinant enzymes are transported into the extracellular medium or into certain compartments of the host cells. Both purification and investigation of the biological activity of the enzymes may consequently be facilitated.

When expressing the enzymes according to the invention, it may prove convenient to modify individual codons. Purposeful replacement of bases in the coding region may here also be advisable if the codons used in the

parasites differ from the codon use in the heterologous expression system, in order to ensure optimal synthesis of the protein.

The enzymes according to the invention may furthermore be obtained under standardised conditions by *in vitro* translation by methods known to the person skilled in the art. Systems suitable for this purpose are rabbit reticulocyte and wheat germ extracts and bacterial lysates. *In vitro* transcribed mRNA may also be translated into *Xenopus* oocytes.

Oligo- and polypeptides, the sequences of which are derived from the peptide sequence of the enzymes according to the invention, may be obtained by chemical synthesis. Given appropriate selection of the sequences, such peptides have properties which are characteristic of the enzymes according to the invention. Such peptides may be produced in large quantities and are particularly suitable for investigating the kinetics of enzyme activity, regulation of enzyme activity, the three-dimensional structure of the enzymes, inhibition of enzyme activity by various chemicals and pharmaceuticals and the binding geometry and binding affinity of various ligands.

DNA with the nucleotides from sequences SEQ ID no. 1, 3 and 5 are preferably used for the recombinant production of the enzymes according to the invention.

The invention accordingly moreover relates to a process for screening for compounds which inhibit the deoxy-D-xylulose phosphate metabolic pathway. According to this

process, a host organism, which contains a recombinant expression vector, wherein the vector comprises at least a portion of the oligonucleotide sequence according to SEQ ID no. 1, SEQ ID no. 3 or SEQ ID no. 5 or variants or homologues thereof, is provided, as is a compound which is suspected to have antimicrobial, antiparasitic, antibacterial, antiviral and antimycotic action in humans and animals or an antimicrobial, antiviral, bactericidal, herbicidal or fungicidal activity in plants. The host organism is then brought into contact with the compound and the activity of the compound determined.

The present invention also provides methods for determining the enzymatic activity of the gcpE protein. Said activity may be determined using known methods. Determination is performed by detecting the phosphorylation of a sugar or of a phosphorus sugar or of a precursor of isoprenoid biosynthesis, in particular the phosphorylation of 2-C-methyl-D-erythritol, 2-C-methyl-D-erytritol phosphate, in particular 2-C-methyl-D-erythritol 4-phosphate, 2-C-methyl-D-erythrose, 2-C-methyl-D-erythrose phosphate, in particular 2-C-methyl-D-erythrose 4-phosphate. The present invention also provides the use of this measurement method for identifying substances which inhibit the activity of the particular enzymes.

The enzymatic activity of DOXP synthase and DOXP reductoisomerase may be detected in a single step by determining the conversion of glyceraldehyde 3-phosphate into 2-C-methylerythritol 4-phosphate.

Determination of the activities of DOXP synthase and DOXP reductoisomerase proceeds analogously. Fluorimetric methods described by Querol *et al.* are also suitable for determining DOXP synthase activity (Querol *et al.*, abstracts, 4th European Symposium on Plant Isoprenoids, Barcelona, 21-23 April 1999).

Claims

1. DNA sequences which code for a polypeptide with the amino acid sequence shown in SEQ ID no. 2 or for an analogue or derivative of the polypeptide according to SEQ ID no. 2, in which one or more amino acids have been deleted, added or replaced by other amino acids.

2. DNA sequences which code for a polypeptide with the amino acid sequence shown in SEQ ID no. 4 or for an analogue or derivative of the polypeptide according to SEQ ID no. 4, in which one or more amino acids have been deleted, added or replaced by other amino acids.

3. DNA sequences which code for a polypeptide with the amino acid sequence shown in SEQ ID no. 6 or for an analogue or derivative of the polypeptide according to SEQ ID no. 6, in which one or more amino acids have been deleted, added or replaced by other amino acids.

4. DNA sequence according to one of claims 1 to 3, characterised in that it also comprises functional regulation signals, in particular promoters, operators, enhancers, ribosomal binding sites.

5. DNA sequence with the following sub-sequences

   i) promoter which is active in viruses, eukaryotes and prokaryotes and ensures the formation of an RNA in the intended target tissue or target cells,

ii) DNA sequences according to one of claims 1 to 3,

iii) 3' untranslated sequence which, in viruses, eukaryotes and prokaryotes, results in the addition of poly(A) residues onto the 3' end of the RNA.

6. Process for the production of transgenic viruses, eukaryotes and prokaryotes for modifying the isoprenoid content, characterised in that a DNA sequence according to claim 4 or 5 is transferred and incorporated into the genome of viruses, eukaryotic and prokaryotic cells with or without use of a vector.

7. Transgenic systems, in particular plants and plant cells which contain one or more DNA sequences according to claims 1 to 5 as "foreign" or "additional" DNA, which sequences are expressed.

8. Expression vector containing one or more DNA sequences according to claims 1 to 5.

9. Protein which is involved in the 1-deoxy-D-xylulose 5-phosphate metabolic pathway and a) is coded by DNA sequences SEQ ID no. 1, 3 or 5 or b) is coded by DNA sequences which hybridise with DNA sequences SEQ ID no. 1, 3, 5 or fragments of these DNA sequences in the DNA region which codes for the mature protein.

10. Protein according to claim 9, obtainable from the culture supernatants of parasites or from the

disrupted parasites and purification by chromatographic and electrophoretic methods.

11. Protein according to one of claims 9 and 10, characterised in that it a) is the product of viral, prokaryotic or eukaryotic expression of exogenous DNA, b) is coded by sequences SEQ ID no. 1, 3 or 5 or is coded by DNA sequences which hybridise with DNA sequences SEQ ID no. 1, 3, 5 or fragments of these DNA sequences in the DNA region which codes for the mature protein, or c) is coded by DNA sequences which would hybridise without degeneration of the genetic code with the sequences defined in b) and which code for a polypeptide with a corresponding amino acid sequence.

12. Protein according to one of the preceeding claims characterised in that it comprises the amino acid sequences SEQ ID no. 2, 4 or 6.

13. Process for determining the enzymatic activity of the gcpE protein, characterised in that phosphorylation of a sugar or of a phosphorus sugar or of a precursor of isoprenoid biosynthesis, in particular the phosphorylation of 2-C-methyl-D-erythritol, 2-C-methyl-D-erytritol phosphate, in particular 2-C-methyl-D-erythritol 4-phosphate, 2-C-methyl-D-erythrose, 2-C-methyl-D-erythrose phosphate, in particular 2-C-methyl-D-erythrose 4-phosphate, and of phosphate and alcohol precursors, is detected.

14. Process according to claim 13, characterised in that phosphorylation of the following phosphates or alcohols is detected:

$CH_2(OH)-C(CH_3)=C(OH)-CH_2-O-PO(OH)_2$,

$CH_2(OH)-C(CH_3)=C(OH)-CH_2-OH$,

$CH_2(OH)-CH(CH_3)-CO-CH_2-O-PO(OH)_2$,

$CH_2(OH)-CH(CH_3)-CO-CH_2OH$

$CH_2=C(CH_3)-CO-CH_2-O-PO(OH)_2$,

$CH_2=C(CH_3)-CO-CH_2-OH$,

$CH_2=C(CH_3)-CH(OH)-CH_2-O-PO(OH)_2$,

$CH_2=C(CH_3)-CH(OH)-CH_2-OH$,

$CH_2(OH)-C(=CH_2)-C(OH)-CH_2-O-PO(OH)_2$,

$CH_2(OH)-C(=CH_2)-C(OH)-CH_2-OH$

$CHO-CH(CH_3)-CH(OH)-CH_2-O-PO-(OH)_2$,

$CHO-CH(CH_3)-CH(OH)-CH_2-OH$,

$CH_2(OH)-C(OH)(CH_3)-CH=CH-O-PO(OH)_2$,

$CH_2(OH)-C(OH)(CH_3)-CH=CH-OH$

$CH(OH)=C(CH_3)-CH(OH)-CH_2-O-PO(OH)_2$,

$CH(OH)=C(CH_3)-CH(OH)-CH_2-OH$,

$(CH_3)_2HC-CO-CH_2-O-PO(OH)_2$,

$(CH_3)_2HC-CO-CH_2-O-H$,

$(CH_3)_2HC-CH(OH)-CH_2-O-PO(OH)_2$,

$(CH_3)_2HC-CH(OH)-CH_2-O-H$.

15. Process for the combined determination of the enzymatic activity of DOXP synthase and of DOXP reductase, characterised in that the conversion of glyceraldehyde 3-phosphate into 2-C-methylerythritol 4-phosphate is detected.

16. Process for screening a compound for the treatment of infectious processes in humans and animals, wherein the process comprises:

a) provision of a host cell which contains a recombinant expression vector, wherein the vector comprises at least a portion of the oligonucleotide sequence according to SEQ ID no. 1, SEQ ID no. 3 or SEQ ID no. 5 or variants or analogues thereof, and moreover of a compound suspected to have antimycotic, antibiotic, antiparasitic or antiviral action in humans and animals,

b) bringing the host cell into contact with the compound and

c) determining the antimicrobial, antimycotic, antibiotic, antiparasitic or antiviral action of the compound.

17. Process for screening for compounds for treating plants, wherein the process comprises:

a) provision of a host cell which contains a recombinant expression vector, wherein the vector comprises at least a portion of the oligonucleotide sequence according to SEQ ID no. 1, SEQ ID no. 3 or SEQ ID no. 5 or variants or analogues thereof, and moreover of a compound suspected to have antimicrobial, antiviral, antiparasitic, bactericidal, fungicidal or herbicidal action in plants,

b) bringing the host cell into contact with the compound and

c) determining the antimicrobial, antiviral, antiparasitic, bactericidal, fungicidal or herbicidal action of the compound.

18. Use of DNA according to one of claims 1 tc 5 or of proteins according one of claims 9 to 12 or of transgenic systems according to claim 7 for the prevention or treatment of diseases in humans and animals.